

# Survey Paper on Multilingual Live Call Translation Using Deep Learning

Vishal Tyagi<sup>1</sup>, Divyansh Agrawal<sup>1</sup>, Vibhu Sehgal<sup>1</sup>, Nishant Shokeen<sup>1</sup>, Nishi Gupta<sup>2,\*</sup>

## Abstract

*This research work surveys cutting-edge language translation technologies, including multi-lingual, real-time translation, voice recognition, speech-to-text conversion, and transcription in the hearing process. The study explores the complex mechanisms behind voice call language translation, focusing on sophisticated machine learning models integrated with cloud-based or local applications to facilitate seamless communication across language barriers. Furthermore, conducting research in live communication analyzes the complexity of text and voice techniques to deliver translated content in timely written and audio formats. Through a comprehensive analysis of current developments, challenges, and future possibilities, this study is analytically valuable to researchers, practitioners, and enthusiasts. Significant advances in natural language processing and machine learning have been shown, and by including advanced methods with deep learning and neural networks, along with strengthening learning, the research aims to stimulate innovation and further development in this dynamic field.*

**Keywords:** Language translation, deep learning, machine learning models, natural language processing, real-time translation, speech-to-text conversion, voice recognition

## INTRODUCTION

In a world where global communication is the cornerstone of progress and collaboration, the ability to seamlessly bridge language barriers in real-time interactions has emerged as a critical need. The fusion of Artificial Intelligence (AI) and Machine Learning (ML) technologies has revolutionized the landscape of live call translation, offering a transformative solution to enable fluid multilingual conversations across diverse contexts. This study delves into the realm of live call translation, exploring the innovative applications of AI and ML in breaking down linguistic barriers and fostering cross-cultural communication.

### \*Author for Correspondence

Nishi Gupta  
E-mail: nishigupta@ncuindia.edu

<sup>1</sup>Student, Department of Computer Science and Engineering, The NorthCap University, Gurgaon, Haryana, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, The NorthCap University, Gurgaon, Haryana, India

Received Date: April 09, 2024

Accepted Date: June 18, 2024

Published Date: June 29, 2024

**Citation:** Vishal Tyagi, Divyansh Agrawal, Vibhu Sehgal, Nishant Shokeen, Nishi Gupta. Survey Paper on Multilingual Live Call Translation Using Deep Learning. Journal of Image Processing & Pattern Recognition Progress. 2024; 11(2): 13–21p.

The constructive collaboration between AI and ML has catalyzed the development of sophisticated speech-to-speech translation systems, heralding a new era where language diversity no longer impedes effective communication. By leveraging advanced neural networks, natural language processing algorithms, and state-of-the-art speech recognition techniques, researchers and technologists are pushing the boundaries of what is achievable in real-time translation scenarios.

Through a comprehensive analysis of existing research, methodologies, and technological advancements, this study aims to unravel the intricacies of live call translation using AI and ML

---

technologies. From addressing the challenges of translating oral languages to exploring the nuances of accuracy and reliability in live call translations, this research endeavors to shed light on the potential and limitations of AI-driven language solutions in fostering global connectivity and understanding. This introduction succinctly introduces the research topic, emphasizes the role of AI and ML in live call translation, and sets the stage for exploring the complexities and potentials of leveraging technology to enhance cross-language communication.

## **BACKGROUND**

The advancement of human civilization has always been accompanied by the challenge of bridging linguistic barriers. Traditional methods of translation, reliant on human interpreters or written translations, have struggled to keep pace with the demands of real-time communication in our interconnected world. However, recent strides in artificial intelligence, machine learning, and natural language processing have revolutionized the field of translation, particularly with the emergence of real-time translation technologies.

Live call translation systems represent a significant breakthrough, offering instantaneous interpretation of spoken language. Powered by sophisticated algorithms and neural networks, these systems decode speech in real-time, enabling seamless communication across linguistic boundaries. This innovation not only enhances cross-cultural dialogue but also promotes inclusivity and accessibility by breaking down language barriers on the fly.

Despite their transformative potential, implementing live call translations presents challenges such as ensuring accuracy across languages and managing technical complexities like latency and data security. Nonetheless, the benefits of these systems are manifold, particularly in domains such as business, healthcare, and education, where effective communication is paramount.

This study explores the implementation of live call translations, examining the methodologies, technologies, and challenges involved. By shedding light on the intricacies of real-time translation technology, this study aims to deepen our understanding of its applications and implications in our increasingly globalized world. Through practical insights and analysis, we seek to highlight the transformative impact of live call translations on cross-cultural communication dynamics and global interconnectedness.

## **LITERATURE REVIEW**

The practice of live subtitling, commonly referred to as "respeaking", constitutes a dynamic methodology employed in live call translations. Within this approach, translators undertake the real-time conversion of spoken language into written text to facilitate immediate translation for live broadcasts [1]. This technique is particularly prevalent in scenarios characterized by rapid speech delivery and the need for instantaneous translation, such as sporting events or news broadcasts. The utilization of translation memories represents another essential technique in live call translations. Translation memories are databases that store previously translated texts, serving as repositories of linguistic knowledge and reference materials.

The process of live subtitling involves several key steps to ensure accurate and timely translation. First, skilled interpreters, known as respeakers, listen to the spoken dialogue and repeat it into a specialized speech recognition software, often utilizing voice recognition technology or stenographic keyboards [2]. The software then transcribes the spoken words into written text in real-time, generating subtitles that are displayed alongside the live broadcast feed. Live subtitling finds application in a diverse range of contexts beyond sporting events and news broadcasts. It is commonly utilized in live television programs, conferences, lectures, and public speeches to enhance accessibility for viewers with hearing impairments or language barriers. Furthermore, live subtitling enables international audiences to access content in their native languages, fostering inclusivity and cross-cultural communication in a globalized world.

One notable feature of live subtitling is its ability to accommodate the nuances of live speech, including variations in tempo, accents, and speech patterns. Respeakers must possess an elevated level of proficiency in both the source and target languages to accurately capture the meaning and intent of the spoken dialogue [3]. Additionally, they may employ techniques such as paraphrasing or summarization to convey complex ideas concisely within the constraints of real-time translation. Despite its effectiveness, live subtitling is not without its challenges. The rapid pace of live broadcasts necessitates quick decision-making and the ability to adapt to unforeseen circumstances, placing considerable pressure on speakers. Moreover, the accuracy of live subtitling may be affected by environmental factors such as background noise or poor audio quality, leading to potential errors in transcription. Additionally, the reliance on speech recognition technology introduces the risk of mistranslations or misinterpretations, particularly for languages with complex phonetic structures or dialectal variations.

In practice, translation memories play a crucial role in streamlining the translation process and enhancing both the speed and accuracy of live translations. When translators encounter segments of text that have been previously translated and stored in the database, the translation memory system automatically retrieves and suggests the corresponding translation. This functionality not only accelerates the translation process by eliminating the need for redundant translation efforts but also ensures consistency in terminology and style across different translations [4]. Moreover, translation memories can be customized and tailored to specific projects or domains, allowing translators to leverage domain-specific terminology and pre-approved translations. This ensures that translations align with the linguistic and stylistic preferences of the target audience, thereby improving the overall quality and readability of the translated content.

Despite their numerous benefits, translation memories are not without limitations. One common challenge is the need for continuous maintenance and updating to ensure the relevance and accuracy of stored translations over time. Additionally, translation memories may not always provide suitable matches for every segment of text, particularly in the case of highly specialized or context-dependent content. In such scenarios, translators may need to rely on their linguistic expertise and judgment to produce accurate and contextually appropriate translations [5].

In contemporary research, the Simultaneous Speech Translation (SimulST) model has emerged as a focal point for its pioneering approach to commencing translation before the entirety of the audio input becomes available. This entails making decisions at each time regarding whether to process the latest information or utilize existing data. Within this framework, decision policies play a crucial role, with fixed strategies like the wait-k policy gaining traction due to their simplicity and efficacy [6]. The SimulST architecture operates on a direct model, featuring both an audio encoder and a text decoder. This configuration affords the advantage of lower latency compared to cascade systems, which involve separate speech recognition and translation stages. Particularly noteworthy is the adaptation of SimulST systems for live subtitling applications, where they incorporate human-driven subtitle segmentation decisions. Experimentation with various display modes, such as word-for-word, blocks, and scrolling lines, aims to optimize reading speed, delay, and subtitle quality. Examining the literature reveals intricate details regarding the training settings of SimulST models. These settings typically involve transformer-based architectures with specific configurations for encoder and decoder layers. During training, the models are optimized using cross-entropy loss and the Adam optimizer. Additionally, a pre-decision module is employed to determine READ/WRITE actions based on token-level information.

Despite the advancements made in SimulST technology, certain gaps and drawbacks persist. One notable limitation is the inherent complexity of decision-making in real-time translation scenarios, which can lead to errors or suboptimal performance. Furthermore, the effectiveness of fixed decision policies like the wait-k strategy may diminish in contexts with variable input characteristics or language

patterns. Additionally, the reliance on token-level information for pre-decision making may overlook broader contextual cues, potentially resulting in inaccuracies or inconsistencies in translation output. These challenges underscore the need for further research to address the limitations of SimulST models and enhance their robustness and adaptability in diverse real-world settings. The MeetDot system is a sophisticated videoconferencing platform developed to enhance communication among multilingual participants by providing live translation captions in English, Chinese, Spanish, and Portuguese. It integrates automatic speech recognition (ASR) and machine translation (MT) technologies in a cascade to deliver real-time translations with features like smooth scrolling captions and a word-guessing game for evaluation. The system architecture includes a frontend for users and a backend with ASR and MT modules, allowing flexibility in service integration. MeetDot focuses on optimizing key metrics such as accuracy, latency, and caption stability, offering an innovative cross-lingual word-guessing game for performance evaluation. The system's modular design enables the use of different ASR and MT services, enhancing its adaptability and performance. Additionally, the system implements strategies to reduce caption flicker and enhance user experience, such as biased decoding during MT, smooth scrolling captions, and preserving line breaks to prevent words from jumping back and forth. The system's evaluation includes intrinsic metrics like accuracy, latency, and erasure, along with an extrinsic evaluation through a cross-lingual word-guessing game. The MeetDot system is planned to be released as open-source software to support further research in this area [7].

Neural Machine Translation (NMT) is a potent tool for achieving precise translation, with ongoing research focusing on its adaptation for real-time language conversion during phone calls, promising more accurate and natural-sounding interpretations. This approach harnesses neural networks to comprehend intricate language structures and generate translations, with studies exploring various architectures and training methods to enhance accuracy and speed. Efforts are also directed towards developing multilingual NMT models capable of handling multiple languages simultaneously, necessitating innovative techniques for cross-language information sharing. However, integrating NMT for real-time processing during calls presents unique challenges, prompting investigations into model optimization and efficient inference methods. Despite being a work in progress, multilingual NMT models demonstrate potential in simultaneous language translation. With continued advancement, NMT holds the potential to revolutionize communication by enabling seamless real-time translations during phone conversations, facilitating effective cross-linguistic interaction [8].

Context-aware translation models represent an innovative approach that holds immense promise for enhancing the accuracy and naturalness of translations in real-time communication scenarios. These models leverage advanced techniques such as neural machine translation with attention mechanisms to adjust translation output based on contextual information dynamically [9]. By incorporating contextual cues such as the surrounding dialogue, speaker intent, and topic continuity, context-aware translation models can generate translations that are linguistically accurate but contextually appropriate. Research in this area has demonstrated significant advancements in improving translation quality by ensuring that translations remain aligned with the evolving conversation. Studies have explored various strategies for integrating contextual information into translation models, including the use of context vectors, dynamic attention mechanisms, and context-aware decoding algorithms. Moreover, research efforts have focused on developing robust evaluation metrics to assess the performance of context-aware translation models in real-world settings [9]. Overall, the growing body of research on context-aware translation models underscores their potential to revolutionize live call translations by enabling translators to produce more accurate, natural, and contextually relevant translations in real-time communication scenarios.

Consecutive interpretation offers a deliberate approach for context aware translation, allowing translators to meticulously process and analyze the source content before delivering the translation in the target language [9]. While this method may slow down the pace of conversation, it enables translators to ensure accuracy and clarity in their translations. By listening to segments of the source

speech and then delivering the translation, translators can consider contextual nuances and convey the intended meaning accurately. Consecutive interpretation is particularly effective in situations where the content is complex or nuanced, as it allows translators to provide accurate and contextually appropriate translations.

Speaker diarization is a pivotal technique in the domain of live call translations, offering a systematic approach to segmenting and identifying speakers within a conversation or audio recording. This methodology has garnered significant attention in literature due to its instrumental role in enhancing the accuracy and contextual relevance of translated content [10]. In the context of live call translations, speaker diarization systems are employed to automatically partition spoken discourse into distinct segments corresponding to individual speakers, thereby facilitating more precise and contextually relevant translations [10]. Research in speaker diarization has highlighted its importance in addressing challenges associated with multi-speaker scenarios, such as group discussions, conference calls, or interviews. Studies have explored various approaches to speaker diarization, including clustering techniques, deep learning models, and hybrid systems combining both statistical and neural network-based methodologies. These approaches aim to accurately identify speakers and segment their speech, enabling translators to attribute dialogue to specific speakers and tailor translations accordingly.

Moreover, the integration of speaker diarization into live call translation systems has demonstrated notable improvements in translation accuracy and coherence. By providing translators with a clear delineation of speakers and their respective speech segments, speaker diarization enhances the ability to capture speaker-specific nuances, intonations, and speech patterns, thus facilitating more nuanced and contextually appropriate translations [10]. Furthermore, the literature underscores the versatility of speaker diarization beyond traditional applications, with studies exploring its potential in specialized domains such as call center analytics, forensic analysis, and speech recognition. These investigations highlight the robustness and adaptability of speaker diarization techniques across diverse domains and underscore their broader implications for advancing research in speech processing and human-computer interaction.

Automatic evaluation metrics are commonly used to assess the quality of machine translation and speech recognition systems employed in live call translation. While automatic metrics provide quantitative evaluations, human evaluation is crucial for assessing the quality, naturalness, and adequacy of live call translations. Techniques like fluency/adequacy scoring, ranking translations, error annotation, and post-editing are commonly used for human evaluation [11, 12]. Human evaluation helps capture aspects like context, pragmatics, and cultural nuances that automatic metrics may miss.

Brown *et al.* highlight the multifaceted nature of the challenges like facilitating seamless exchange of information across linguistic barriers, ensuring accurate and reliable translation during live calls, encompassing linguistic nuances, cultural contexts, and technical constraints [13]. They emphasize the importance of translation accuracy, fluency, and cultural sensitivity in maintaining effective communication. Hazelwood *et al.* further delve into the technical constraints, such as latency and bandwidth limitations, which can impact translation quality [14]. Advancements in hardware and networking infrastructure that contribute to addressing these technical challenges.

To optimize translation quality, researchers have explored various strategies and technologies. Gracia *et al.* analyzed automated quality assessment algorithms for real-time monitoring [15], while Ribeiro *et al.* advocate for human-in-the-loop approaches for error correction and refinement [16]. Wang *et al.* emphasize the importance of continuous training and evaluation of translation models to adapt to evolving linguistic patterns [17]. Additionally, advancements in neural machine translation models [18], Natural Language Understanding systems [19], and speech recognition technologies [20] have significantly contributed to enhancing translation accuracy and fluency.

Quality assurance protocols play a crucial role in ensuring consistent and high-quality translations. Callison-Burch *et al.* discuss the development of standardized evaluation metrics for translation quality [21], while Bahdanau *et al.* advocate for the establishment of feedback loops and user feedback mechanisms [22]. Through case studies and real-world implementations, researchers have evaluated the efficacy of these quality control frameworks. For instance, Yang *et al.* analyzed quality control frameworks in live call translation platforms, while it examines their impact on user satisfaction levels [23].

Looking ahead, researchers acknowledge the need for continued advancements in quality control mechanisms [24]. Ethical considerations in AI-driven solutions for translation, highlighting the importance of transparency and accountability. Additionally, scalability challenges and the need for standardized evaluation metrics remain areas of focus for future research [25].

Ogundokun *et al.* developed an android-based language translation application that enables translating between the most widely spoken global languages like English, Spanish, French, Hindi, Arabic, and Chinese [26]. The application utilizes Google's real-time translation API and natural language processing techniques implemented using Java. The system allows users to input text or capture images containing text, which are then translated to the selected target language using IBM Watson's Language Translator service running on the cloud [26]. The application was evaluated for accuracy, translation time, memory usage, and battery impact, demonstrating promising performance.

## **METRICS FOR MULTILINGUAL MODELS' EVALUATION**

### **BLEU (Bilingual Evaluation Understudy)**

BLEU compares machine-translated text against one or more reference human translations and calculates the geometric mean of n-gram precision scores [27]. Despite its widespread use, BLEU has been criticized for its limitations in capturing semantic equivalence and adequacy [28].

### **TER (Translation Edit Rate)**

TER measures the number of edits (insertions, deletions, substitutions, and shifts) required to transform the machine translation output into an acceptable reference translation [29]. Lower TER scores indicate better translation quality, and the metric has been found to correlate well with human judgments of translation quality [30].

### **METEOR (Metric for Evaluation of Translation with Explicit Ordering)**

METEOR considers not only the overlap of words but also their ordering and matching of phrases [31]. It combines several features, including precision, recall, and fragmentation penalties, and has been shown to correlate better with human judgments than other metrics like BLEU [32].

### **Word Error Rate (WER)**

WER is a widely used metric for evaluating the performance of Automatic Speech Recognition (ASR) systems employed in live call translation [33]. It measures the editing distance between the hypothesized text (ASR output) and the reference transcription, considering insertions, deletions, and substitutions. Lower WER values indicate better speech recognition accuracy.

## **DRAWBACKS OF EXISTING TECHNIQUES**

The extensive body of research on live subtitling, translation memories, and emerging technologies for real-time translation unveils several areas ripe for further exploration and refinement. While existing studies have diligently examined the efficacy of live subtitling and translation memories in widely accessible contexts such as news broadcasts and conferences, there remains a noticeable gap in understanding their applicability in more specialized domains like legal proceedings or medical consultations. Exploring these niche areas could provide invaluable insights into tailoring translation solutions to meet specific communication needs and challenges.

Moreover, the ongoing development of Simultaneous Speech Translation (SimulST) models represents a promising frontier in real-time translation technology. However, to fully realize their potential, it is imperative to enhance their robustness and adaptability. This entails the refinement of decision-making algorithms to navigate the complexities of real-world translation scenarios and the integration of broader contextual cues beyond token-level information. By incorporating a deeper understanding of context and discourse dynamics, SimulST models can produce more accurate and contextually relevant translations, particularly in dynamic conversational settings.

In parallel, the optimization of user experience and interface design in platforms like the MeetDot system is critical for fostering seamless interaction and effective communication. As real-time translation technologies become increasingly integrated into daily communication channels, prioritizing user-centric design principles can enhance accessibility and usability, ultimately improving the overall efficacy of cross-lingual communication.

Finally, as real-time translation technologies continue to evolve, it is essential to consider the ethical and social implications associated with their deployment. This includes ensuring transparency, accountability, and fairness in the development and implementation of translation systems, as well as addressing issues of data privacy and equity in access to translation services across diverse linguistic communities. By proactively addressing these ethical considerations, researchers and practitioners can mitigate potential risks and maximize the positive impact of real-time translation technologies on global communication and understanding.

## CONCLUSION

The study presents a comprehensive overview of methodologies and technologies employed in live call translations, highlighting their significance, challenges, and potential for advancement. The practice of live subtitling, or "respeaking", and the utilization of translation memories emerge as fundamental techniques, offering solutions for real-time translation needs in various contexts. Live subtitling, with its ability to accommodate rapid speech delivery and its applications beyond broadcasting, demonstrates effectiveness in enhancing accessibility and fostering cross-cultural communication. However, challenges such as rapid decision-making and reliance on speech recognition technology underscore the need for continuous improvement.

The introduction of context-aware translation models represents a significant advancement in addressing these challenges. By dynamically adjusting translation output based on contextual cues, these models promise to revolutionize live call translations by ensuring accuracy, naturalness, and context relevance. Additionally, consecutive interpretation offers a deliberate approach, enabling translators to provide accurate and contextually appropriate translations, especially in complex or nuanced scenarios.

Furthermore, the Simultaneous Speech Translation (SimulST) model and the MeetDot system highlight innovative technologies aimed at improving real-time translation accuracy, latency, and user experience. These advancements, coupled with ongoing research in areas such as Neural Machine Translation (NMT) and multilingual models, signal promising prospects for the future of live call translations.

Overall, the survey underscores the importance of adapting methodologies and technologies to meet the evolving demands of live call translations while addressing inherent challenges. By leveraging advancements in technology and incorporating insights from interdisciplinary research, the field is poised to enhance the quality, efficiency, and accessibility of real-time translation services, ultimately fostering effective communication across linguistic boundaries.

---

**REFERENCES**

1. Gambier Yves. Translations| Rapid and Radical Changes in Translation and Translation Studies. *Int J Commun.* 2016; 10: 887–906.
2. Remael Aline, Gert Vercauteren. The translation of recorded audio description from English into Dutch. *Perspect: Stud Transl.* 2010; 18(3): 155–171.
3. Szarkowska Agnieszka, et al. Cognitive load in intralingual and interlingual respoking—a preliminary study. *Pozn Stud Contemp Linguist.* 2016; 52(2): 209–233.
4. Hutchins John. Machine translation: A concise history. *Journal of Translation Studies* 13 (1 & 2) (2010), 29–70.
5. Sager Juan C. A practical course in terminology processing. *A Practical Course in Terminology Processing.* Amsterdam; Philadelphia: J. Benjamins Pub. Co.; 1990; 1–270.
6. Turchi M, Fantinuoli C. Proceedings of the 1st Workshop on Automatic Spoken Language Translation in Real-World Settings (ASLTRW). In *Proceedings of the 1st Workshop on Automatic Spoken Language Translation in Real-World Settings (ASLTRW)*. 2021 Aug.
7. Arkhangorodsky Arkady, et al. MeetDot: Videoconferencing with live translation captions. *arXiv preprint arXiv:2109.09577.* 2021.
8. Wu Yonghui, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144.* 2016.
9. Lu Ziyao, et al. Exploring multi-stage information interactions for multi-source neural machine translation. *IEEE/ACM Trans Audio Speech Lang Process.* 2021; 30: 562–570.
10. Anguera Xavier, et al. Speaker diarization: A review of recent research. *IEEE Trans Audio Speech Lang Process.* 2012; 20(2): 356–370.
11. Callison-Burch C, Fordyce C, Koehn P, Monz C, Schroeder J. (Meta-) evaluation of machine translation. *Proceedings of the Second Workshop on Statistical Machine Translation.* 2007; 136–158.
12. Graham Y, Baldwin T, Moffat A, Zobel J. Continuous measurement scales in human evaluation of machine translation. *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse.* 2013; 33–41.
13. Brown Tom, et al. Language models are few-shot learners. *33th Int Conf on Advances in neural information processing systems.* 2020; 1877–1901.
14. Hazelwood Kim, et al. Applied machine learning at facebook: A datacenter infrastructure perspective. *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA).* 2018; 620–629.
15. Gracia Jorge, Iian Kernerman, Besim Kabashi. Results of the translation inference across dictionaries 2021 shared task. *CEUR workshop proc. No. ART-2021-131934.* 2021; 208–220.
16. Ribeiro Marco Tulio, Sameer Singh, Carlos Guestrin. "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* 2016; 1135–1144.
17. Wang Xing, et al. Neural machine translation advised by statistical machine translation. *31st Proceedings of the AAAI conference on artificial intelligence.* 2017; 3330–3336.
18. Luong Minh-Thang, Hieu Pham, Manning Christopher D. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025.* 2015.
19. Vaswani Ashish, et al. Attention is all you need. *30th Int Conf on Advances in neural information processing systems.* 2017; 6000–6010.
20. Chorowski Jan K, et al. Attention-based models for speech recognition. *28th Int Conf on Advances in neural information processing systems.* 2015; 1: 577–585.
21. Song X, Cohn T, Specia L. BLEU Deconstructed: Designing a Better MT Evaluation Metric. *Int. J. Comput. Linguistics Appl.* 2013 Jul;4(2):29–44.
22. Bahdanau Dzmitry, Kyunghyun Cho, Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473.* 2014.
23. Yang Yilin, et al. Improving multilingual translation by representation and gradient regularization. *arXiv preprint arXiv:2109.04778.* 2021.

24. Sutskever Ilya, Oriol Vinyals, Le Quoc V. Sequence to sequence learning with neural networks. 27th Int Conf on Advances in neural information processing systems. 2014; 2: 3104–3112.
25. Callison-Burch Chris, et al. Findings of the 2010 joint workshop on statistical machine translation and metrics for machine translation. Proceedings of the Joint 5th Workshop on Statistical Machine Translation and Metrics MATR. 2010; 17–53.
26. Ogundokun RO, Awotunde JB, Misra S, Segun-Owolabi T, Adeniyi EA, Jaglan V. An android based language translator application. J Phys: Conf Ser. 2021; 1767(1): 012032.
27. Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002; 311–318.
28. Callison-Burch C, Osborne M, Koehn P. Re-evaluating the role of BLEU in machine translation research. 11th Conference of the European Chapter of the Association for Computational Linguistics. 2006; 249–256.
29. Snover M, Dorr B, Schwartz R, Micciulla L, Makhoul J. A study of translation edit rate with targeted human annotation. Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. 2006; 223–231.
30. Snover M, Madnani N, Dorr B, Schwartz R. TER-Plus: Paraphrase, semantic, and alignment enhancements to Translation Edit Rate. Mach Transl. 2009; 23(2–3): 117–127.
31. Banerjee S, Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. 2005; 65–72.
32. Denkowski M, Lavie A. Meteor universal: Language specific translation evaluation for any target language. Proceedings of the 9th Workshop on Statistical Machine Translation. 2014; 376–380.
33. Li H, Ma B, Lee KA. Spoken language recognition: from fundamentals to practice. Proc IEEE. 2013 Feb 6; 101(5): 1136–59.